Spatial scale of GIS-derived categorical variables affects their ability to separate sites by community composition

Holt, Emily A.[1*]; McCune, Bruce[1,2] & Neitlich, Peter[3]

[1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331-2902, USA;
[2]E-mail bruce.mccune@science.oregonstate.edu;
[3]National Park Service, Western Arctic National Parklands, 41A Wandling Road, Winthrop, WA 98862, USA;
E-mail peter_neitlich@nps.gov; *Corresponding author; E-mail emilyholt@lifetime.oregonstate.edu

Abstract

Questions: How well do GIS-derived categorical variables (e.g., vegetation, soils, geology, elevation, geography, and physiography) separate plots based on community composition? How does the ability to distinguish plots by community composition vary with spatial scale, specifically number of patch types, patch size and spatial correlation? Both these questions bear on the effective use of stratifying variables in landscape ecology.

Location: Arctic tundra; Bering Land Bridge National Preserve, northwestern Alaska, USA.

Methods: We evaluated the strength of numerous alternative stratifying variables using the multi-response permutation procedure (MRPP). We also created groups based on lichen community composition, using cluster analyses, and evaluated the relationship between these groups and groupings within categorical variables using Mantel tests. Each test represents different measures of community separation, which were then evaluated with respect to each variable's spatial characteristics.

Results: We found each categorical variable derived from GIS separated lichen communities to some degree. Separation success ranged from strong (Alaska Subsections) to weak (Watersheds and Reindeer Ownership). Lichen community groups derived from cluster analysis demonstrated statistically significant relationships with 13 of the 17 categorical variables. Partialling out effects of spatial distance had little effect on these relationships.

Conclusions: Greater number of patch types and larger average patch sizes contribute to optimal success in separating lichen communities; geographic distance did not appear to significantly alter separation success. Group distinctiveness or strength increased with more patch types or groups. Alternatively, congruence between lichen community types derived from cluster analysis and the 17 categorical variables was inversely related to patch size and spatial correlation.

Keywords: Alaska; Lichen; Patch type; Sampling; Stratification.

Abbreviation: MRPP = Multi-response permutation procedure.

Introduction

Do species and community distributions have a strongest pattern at particular scales, providing clues to important drivers and optimal landscape classification? Some have tackled this problem with geostatistical tools using field-derived point data (e.g., Gustafson 1998; Legendre & Fortin 1989; Meisel & Turner 1998), while others have used various indices to characterize patch attributes for categorical maps (e.g., Gustafson 1998; Li & Reynolds 1994). Within these categorical maps, manipulation of grain and extent has enabled exploration of species patterns on the landscape over a range of scales (Wu 2004). Few studies, however, analyze point community data in relation to varying scale of categorical maps (Gustafson 1998).

The important distinction between the two approaches, field-based point data and categorical maps, is practicality. We define GIS-derived categorical data as an off-site method of obtaining information remotely about any location in the landscape. For many regions, these data are inexpensive and easily accessible prior to an investigation to anyone with a GIS. Categorical maps are used in many facets of ecological studies, most notably as means to focus sampling efforts. Alternatively, point data require site visits, which are always time consuming and often expensive, but are rich in detail, specificity and 'ground truth.'

We gathered detailed field measurements of lichen community composition (i.e. point data) and combined them with categorical variables derived from GIS maps. Our primary interest in these categorical variables was not their biological or environmental basis (e.g., dominant vegetation or type of underlying bedrock), but rather how they vary in space. Many ecologists using GIS maps focus on environmental attributes yet are unaware of how spatial variability of these variables influences their data or findings. Each categorical variable we used has unique spatial characteristics, including differences in patch size, number of patch types and spatial cor-

relation. We questioned, therefore, whether these three attributes of spatial scale confounded our evaluation of how well GIS-based categorical variables separate plots with respect to lichen community composition. For example, one variable with many small patches may have the same number of patch types as another variable with only a few large patches. Likewise, two variables may share the same size patches yet have a different overall number of patch types. Our primary purpose of this study was to use GIS-based categorical variables to explore scaling effects on vegetation, using lichens as our example system.

An additional motivation for analyzing point data in relation to categorical variables was to identify potentially useful stratifying variables for future studies. The pervasive use of GIS to identify stratifying principles in every system across all biomes underscores the need to understand how spatial characteristics influence these variables' ability to adequately stratify the landscape. Stratification has been praised as one of the most efficient and precise sampling strategies (Cochran 1977; Goedickemeier et al. 1997; Smartt & Grainger 1974). This method divides a heterogeneous population into more homogeneous units that can reduce both types I and II error and reduce variance of the estimate (Cochran 1977; Kernan et al. 1999). Stratified sampling is pervasive throughout scientific research (e.g. Brus 1994; Jongman et al. 2006; Kernan et al. 1999; Olsen et al. 1999). Within ecology, the advent of GIS and availability of spatial data online provide essentially infinite stratification possibilities. Stratifying ecological sampling is often limited only by questions about which, and how many, variables to use. The choice of stratifying variables, however, often depends on the question of interest, extent of area, and availability of datasets (Goedickemeier et al. 1997; Knollová et al. 2005). Optimizing the stratification clearly depends on how strongly the stratifying variables relate to actual landscape patterns of the response variable (Bond & Devine 1991; Brus 1994). It may also depend on the spatial scaling of the strata.

In a separate study, we described lichen community structure and its relationship to environmental factors in the same system (Holt et al. 2007). From this work, we know that lichen community patterning primarily responds to habitat rockiness, substrate (pH) and topography. So in the current paper, our question shifts from what specific variables control vegetation patterns to how spatial characteristics of variables influences their success in perceiving vegetation patterns. We had two main questions: 1. How well do categorical variables separate plots based on lichen community composition? 2. How does changing a variable's spatial characteristics (including patch size, number of patch types and spatial correlation) affect separation of plots by community composition?

Material and Methods

Study site

The Bering Land Bridge National Preserve is located on the Seward Peninsula in northwestern Alaska (65°14'-66°36' N, 162°44'-167°32' W). Temperatures of the Seward Peninsula (mean July temperatures about 10.6°C) are tempered by the oceanic influence of the surrounding Bering and Chukchi seas. Mean annual precipitation, falling primarily in late summer, is 444 mm in the town of Nome, in the southwest corner of the Peninsula. The bulk of the Preserve comprises moist to wet tussock tundra underlain by continuous permafrost alternating with discontinuous permafrost (Van Patten 1990). Rising south from sea level, the Preserve's southern boundary follows the crest of the Bendeleben Mountains to a height of 1040 m. Adding unique geology to an otherwise metamorphic landscape, dolomite outcrops along the east and western boundaries of the Preserve, and historic volcanic activity sprinkles the north and central portions of the Seward Peninsula (Till & Dumoulin 1994).

The general vegetation types present in northwestern Alaska are *Eriophorum* tussock tundra, *Dryas* fellfield, ericaceous-shrub tundra, *Eriophorum-Carex* wet meadow, and solifluction slopes (Viereck et al. 1992). Common shrubs consist of *Salix* spp., *Betula glandulosa*, *B. nana* and some *Alnus crispa*. The herb layer contains mixed *Eriophorum* spp. and *Carex* spp., *Vaccinium* spp., *Arctostaphylos* spp., *Empetrum nigrum*, *Cassiope tetragona*, *Ledum palustre* var. *decumbens* and *Rubus chamaemorus*. The dominant mosses are *Sphagnum* spp. and *Hylocomium splendens*. The lichen flora is dominated by species of *Cladina*, *Cladonia*, *Cetraria*, *Peltigera* and *Stereocaulon*.

Sampling

We used a two-way stratified random sample. One of our stratifying variables was GIS land cover data (Markon & Wesser 1997). Cover types were formulated from classified Landsat data, which reflect the spectral response of the Earth's surface including vascular plants, bryophytes, rock, water, lichens, etc. We used land cover types to discriminate lichen-poor from lichen-dominated areas within the Preserve. A pilot study indicated that four of the ten Landsat-based cover types had an average lichen cover of at least 10%. These four cover types were the only strata we sampled further. Geographic blocks were the other stratifying variable, used to balance sampling across the area of interest. The central portion of the

Table 1. Descriptions of 17 GIS-based categorical variables.

Categorical variable	Description	Source	No. patch types[1]	Patch size[2]
Imagery-based				
UAF Vegetation	Landsat satellite imagery classification	Alaska Geobotany Center	8 (8)	496.2
Cover Types	Landsat satellite imagery classification	Markon & Wesser 1997	4 (10)	0.007
Map Ecological Types	Local-scale ecosystems derived from physiography, geology, topography and satellite imagery	Jorgenson et al. 2004	10 (11)	0.007
Map Vegetation Types	Aggregation of Map Ecological Types variable based on soils, topography and locality	Jorgenson et al. 2004	7 (7)	0.006
Aggregate Ecological Types	Aggregation of Map Ecological Types variable based on vegetation, hydrology and topography	Jorgenson et al. 2004	5 (5)	0.006
Hierarchical Ecoregions				
Ecoregions	Ecological regions of synthesis of geographic distribution of climate, terrain, soils and vegetation	NPS Alaska	2 (2)	5639.1
Physiography	Division of Ecoregions variable based on terrain	NPS Alaska	8 (9)	1025.3
Alaska Sections	Division of Ecoregions variable based on terrain and location	NPS Alaska	7 (7)	1025.3
Alaska Subsections	Division of AK Sections variable (further division of Ecoregions) based on locality	NPS Alaska	13 (15)	375.9
Substrate				
Soil Series	Soils map at series-level of classification	NPS Alaska	10 (10)	216.7
Soil Units	Soils map and unit-level of classification	NPS Alaska	10 (23)	80.38
Surficial Geology	Digital rendering of two USGS maps (Misc. Geological Investigations (West) I-357 and Misc. Geological Investigations (East) I-357)	NPS Alaska	8 (8)	22.84
Other				
Watersheds	Watershed boundaries delineated by Hydrologic Unit Codes (HUC)	NPS Alaska	4 (4)	779.3
Geographic Blocks	Grid cells lumped together to form roughly equal area blocks in all areas of the Preserve	Author-created	20 (21)	434.4
Reindeer Owner	Regions of Seward Peninsula separated by reindeer permit	NPS Alaska	3 (5)	395.9
Alaska Ecosystems	Adapted from 1973 map of "Major Ecosystems of Alaska," based on regional distribution of vegetation in relation to topography, climate, and hydrology	NPS Alaska[3]	4 (5)	303.3
Elevation Bands	Bands of 250-ft increments based on 60-meter National Elevation Dataset (NED) derived from USGS 30-meter DEMs	NPS Alaska	8 (10)	24.0

[1]Number of patch types containing more than one plot, and number in parentheses is the original number of groups in the sampled portion of the Preserve;
[2]Average patch size among all patch types (km^2);
[3]NPS Alaska = National Park Service Alaska GIS Data Clearinghouse.

Preserve, where we focused our sampling, was overlain with a grid in GIS. Grid cells were then aggregated into ca. 21 roughly equal-area geographic blocks. The blocks were each ca. 400 km^2. Within each geographic block, we randomly located one plot from each of the four cover types using a random point generator in a GIS. Two of the four cover types were present in every block, whereas the other two were completely absent from some blocks. Therefore only 65 strata or possible combinations of the two variables, instead of the expected 84, occurred. To avoid an unbalanced sample, the two infrequent cover types were sampled twice in 13 blocks. In total, data from 78 plots within a total of 65 strata from two stratifying variables were collected.

Sample units were circular fixed-area plots with a 34.7-meter radius. Each macrolichen species encountered was assigned an abundance value: 1 = rare (<3 thalli), 2 = uncommon (4-10 thalli), 3 = common (<1% cover), 4 = abundant (1-5% cover), 5 = prolific (6-25% cover) and 6 = dominant (> 26% cover). Species delimitations are described in Holt et al. (2007). GIS datasets provided a range of stratifying variables reflecting vegetation, soils, geology, elevation, geography and physiography at various spatial resolutions (Table 1). Many of these variables were highly correlated (App. 1), yet their intercorrelations and redundancy were inconsequential to our analysis; as our foremost interest was an array of spatial properties from GIS data easily-accessible by researchers of many fields. To our knowledge, lichen community data were not used to help derive classes for these datasets. These data were obtained from the National Park Service Alaska GIS Data Clearinghouse (http://www.nps.gov/akso/gis/), Jorgenson et al. (2004) and Alaska Geobotany Center (http://www.geobotany. uaf.edu/).

Analyses

We first determined if any plots were multivariate outliers by comparing average community distances between plots. One of the 78 plots, Plot 5P, had average Sørensen distance of 3.7 standard deviations from the grand mean of all distances. This plot was an outlier because its total lichen cover (2.5%) and richness (13 species) was far lower than the average (26.4% and 24.3 species). Therefore, plot 5P, which may represent an error in the Landsat classification, was removed from all analyses. Otherwise, modifications to the community matrix were minimal. The coarse, approximately logarithmic, cover class scale alleviated the need for further transformation.

We used two methods to evaluate efficacy of categorical variables in separating lichen communities. One method evaluated separation success using raw community data, or complete species abundances from all 77 plots. The second method, alternatively, used a clustered form of the same community data (groupings based on community composition). We used these two methods and comparisons therein to identify relationships between GIS-derived categorical variables and to contrast detailed versus general compositional patterns. For our first method of evaluating separation success using raw data, we compared GIS-based groupings of plots (GIS variables served as categorical variables by which to group plots) to actual community composition of each plot. We measured group distinctness, or separation success for each individual categorical variable, using multi-response permutation procedure (MRPP; Mielke 1984). MRPP is a non-parametric technique which compares within-group homogeneity of *a priori* groupings to random expectation. Compositional dissimilarity, measured as Sørensen distance, was averaged within each group then pooled across all groups. Statistical significance of these groupings is evaluated by asymptotic approximation (*p*-value), and the strength or distinctness of each group was evaluated by an *A*-statistic, the chance-corrected within-group homogeneity. $A = 1$ indicates perfectly homogenous groups, while $A = 0$ indicates within-group heterogeneity equal to chance expectation. For ecological data, $A = 0.3$ is rather strong (McCune & Grace 2002).
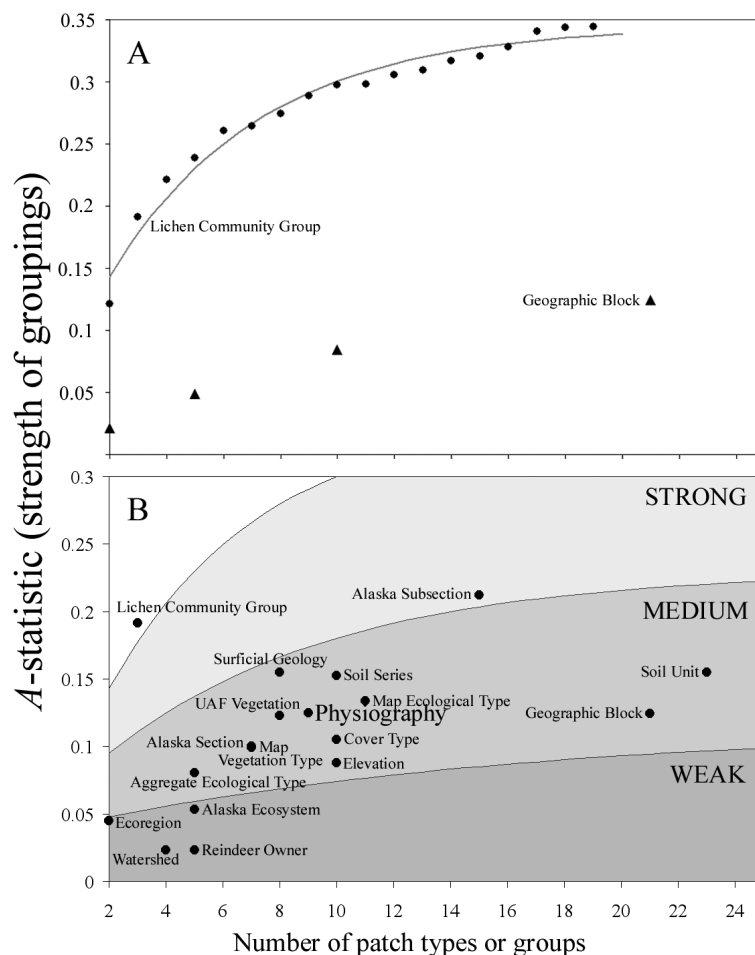

Fig. 1. Scatterplot of number of patch types or groups and *A*-statistics. *A*-statistics tend to increase with more groups or patch types. This relationship is evident with (**A**) two hierarchical variables (lichen community types and geographic blocks) and (**B**) all 17 categorical variables and the lichen community groups. In the top panel (**A**), black triangles represent the geographic blocks variable (original variable contained 21 patch types), which were aggregated to form fewer patch types. The black circles depict 20 and higher-level groupings derived from cluster analysis. Grey line represents non-linear regression of these points ($y = 0.345 - 0.202e^{-0.188\,(x\,-\,2)}$). Since this line is derived directly from lichen community composition, this line is the maximum achievable *A*-statistic at each given number of groups or patch types. We selected three as the optimal number of groups – noted as lichen community groups – because all larger groupings resulted in only minor increases in the *A*-statistic. In the bottom panel (B), the uppermost line represents the same non-linear regression from (A), while the other lines represent families of this curve. The shaded banding between these curves corresponds to levels of success (weak, medium, strong), assessed by the ability to maximize the *A*-statistic while maintaining a manageable number of groupings, the lightest being the optimal level.

For the second method of separation success, we compared GIS-based categorical groupings to clustered groupings that represent a data reduction of community composition itself. The community-based grouping to which we compared each categorical variable was constructed using one-way hierarchical agglomerative cluster analyses in PC-ORD 5 (McCune & Mefford 2005). Plots were combined into groups based on compositional dissimilarity, measured as Sørensen distance. We used flexible beta ($\beta = -0.25$) as the linkage method.

Using cluster analysis, we derived several different groupings of lichen community composition. We pruned the lichen community dendrogram, from 77 plots, to include 20 and higher-level groupings. We chose three as the optimal number of groups (called lichen community groups through the remainder of the paper and in Holt et al. 2007). This solution was chosen because only minor increases in the *A*-statistic in MRPP resulted from solutions with more than three groups and to be consistent with Holt et al. 2007 (Fig. 1a).

To compare our groupings from cluster analysis to groupings from each categorical variable derived from GIS, we used Mantel tests (Mantel 1967) of two symmetrical square matrices of binary values. All matrices were 77×77 plots. Matrix $\mathbf{Y}$ represented differences in cluster groupings. Each element was a binary value indicating whether two plots occurred in the same cluster ($y_{ij} = 0$) or were members of different cluster-based groups ($y_{ij} = 1$). Accordingly, we created 17 $\mathbf{X}$ matrices, one for each categorical variable, again of 77 plots $\times$ 77 plots. Each element indicated when two plots occurred within the same GIS patch type ($x_{ij} = 0$) or different patch types ($x_{ij} = 1$). The diagonals of all matrices held zeros. We then used the standardized Mantel statistic (r) as a measure of congruence, or agreement of group or patch type assignment, between the single $\mathbf{Y}$ and each $\mathbf{X}$.

We also used partial Mantel tests, which controlled for geographic distance between plots, to further understand the results from the community group Mantel test described above. Congruence was estimated between the single $\mathbf{Y}$ cluster-based matrix and each 17 GIS-based categorical $\mathbf{X}$ matrices noted above, while accounting for the linear correlation of a third matrix, a 77 plot by 77 plot geographic distance matrix, $\mathbf{Z}$. Partial Mantel tests were calculated with the method of Smouse et al. (1986) as described by Legendre & Legendre (1998, p. 558).

Finally, we explored the influence of spatial characteristics on each categorical variable's ability to separate plots by lichen community patterns. We used three spatial characteristics; average patch size, total number of unique patch types and spatial correlation. Patch size and number of patch types were calculated in a GIS. The spatial correlation was another set of Mantel tests comparing the geographic distance between plots to group or patch type

assignments determined by each variable. Specifically, matrix $\mathbf{Z}$, a 77×77 geographic distance matrix (in m), was compared to 18 matrices (i.e., 17 GIS-based $\mathbf{X}$s and one cluster-based $\mathbf{Y}$, all described above). The congruence (standardized Mantel statistic, r) between $\mathbf{Z}$ and each $\mathbf{X}$ or $\mathbf{Y}$ represented our spatial correlation.

We then sought relationships between these three spatial characteristics (patch size, number of patch types and spatial correlation) and our two measures of plot separation by lichen community patterns (MRPP of raw community data and community Mantel tests between community-based groups and each categorical variable). We evaluated both linear and non-linear relationships between spatial characteristics and measures of separation, but found linear relationships to suffice.

Results and Discussion

Separation using raw community data

We expected that GIS-derived categorical variables based on habitat and landscape would create the strongest groups to best separate macrolichen species composition. We found, indeed, that all 17 categorical variables separated lichen communities using raw community data more strongly than by chance alone (Table 2). However, the effect size of this difference, or the *A*-statistic, showed some grouping variables were much stronger than others. The largest *A*-statistic was 0.21 for Alaska Subsections, while other variables with fairly high *A*-statistics included lichen community groups and the three substrate variables (Table 2). The two variables used in our sampling stratification, however, cover types and geographic blocks, showed only mid-range *A*-statistics ($A = 0.11$ and 0.12, respectively).

Separation using clustered community group data

We found that 13 of the 17 groupings within the categorical variables were significantly ($\alpha = 0.05$) related to these lichen community groups (Table 2). The strength of these relationships, however, was only moderate to weak (community group Mantel's $r \leq 0.28$). The strongest congruence involved imagery-based variables with small patch sizes (e.g., Aggregate Ecological Types, Map Ecological Types, Map Vegetation Types and Cover Types). Alternatively, variables as weak as or weaker than randomizations of one of the matrices included Reindeer Owner, Ecoregions, Alaska Sections and Watersheds – all coarse groupings reflecting broad landscape patterns less strongly related to vegetation.

Table 2. Results from MRPP analyses (evaluating differences in lichen community composition among grouping variables), spatial correlation (testing congruence of geographic distances and patch type or group assignment), community group Mantel tests (testing congruence of groups from 17 categorical variables to groupings from cluster analysis), and community group Mantel tests controlling for geography (testing congruence of groups from 17 GIS variables to groupings from cluster analysis controlling for geographic distance between plots).

Grouping variable	MRPP[1]	Spatial correlation		Lichen Community Groups			
	A	Mantel's r	p	Mantel's r	p	Partial Mantel's r	P
Lichen Community Groups[2]	0.19	0.03	0.17	NA	NA	NA	NA
Imagery-based Categorical							
UAF Vegetation	0.12	0.22	< 0.01	0.14	< 0.01	0.13	< 0.01
Cover Types	0.11	0.04	0.04	0.24	< 0.01	0.24	< 0.01
Map Ecological Types	0.13	~ 0.00	0.50	0.26	< 0.01	0.26	< 0.01
Map Vegetation Types	0.10	-0.01	0.42	0.26	< 0.01	0.26	< 0.01
Aggregate Ecological Types	0.08	~ 0.00	0.38	0.28	< 0.01	0.28	< 0.01
Hierarchical Ecoregions Categorical							
Ecoregions	0.04	0.52	< 0.01	-0.03	0.23	-0.06	0.02
Physiography	0.12	0.28	< 0.01	0.07	0.02	0.07	0.02
Alaska Sections	0.10	0.39	< 0.01	~ 0.00	0.55	-0.02	0.32
Alaska Subsections	0.21	0.29	< 0.01	0.16	< 0.01	0.15	< 0.01
Substrate Categorical							
Soil Series	0.15	0.20	< 0.01	0.08	0.01	0.08	0.01
Soil Units	0.15	0.28	< 0.01	0.07	0.01	0.07	0.01
Surficial Geology	0.15	0.20	< 0.01	0.19	< 0.01	0.19	< 0.01
Other Categorical							
Watersheds	0.02	0.39	< 0.01	~ 0.00	0.56	-0.02	0.30
Geographic Blocks	0.12	0.30	< 0.01	0.07	< 0.01	0.06	< 0.01
Reindeer Owner	0.02	0.58	< 0.01	-0.04	0.14	-0.07	< 0.01
Alaska Ecosystems	0.05	0.28	< 0.01	0.08	0.02	0.07	0.03
Elevation Bands	0.09	0.21	< 0.01	0.13	< 0.01	0.12	< 0.01

[1] All comparisons among groups using MRPP had $p < 0.005$;
[2] Three-group solution derived from cluster analysis.

Spatial characteristics

We found varying relationships between our three spatial characteristics (number of patch types, patch size and spatial correlation) and two measures of plot separation (MRPP of raw community data and community group Mantel tests). First, spatial correlation was negatively related to both measures of separation success (Fig. 2). We found variables with low spatial correlation strongly separated lichen community groups ($r = -0.95$ of spatial correlation and community group Mantel's r; Fig. 2a). Albeit a weaker relationship, low spatial correlation also related to high separation success with the raw community dataset ($r = -0.45$ of spatial correlation and A-statistics; Fig. 2b). The patterns are strikingly similar excepting the four imagery-based variable with small patches, and removal of these variables would markedly strengthened this correlation ($r = -0.72$).

These indirect relationships suggest that our sample of lichen community patterning is relatively spatial independent, contrary to the generality of autocorrelation often noted in ecological literature (e.g., Legendre & Fortin 1989; Legendre 1993). To further support these findings, we found that partial Mantel tests, factoring out geographic distance, differed only slightly from the regular Mantel tests (Table 2). Spatial proximity of plots, therefore, contributed little or nothing to our ability to separate lichen communities at the scale we sampled. Furthermore, we found spatial correlation was strongly correlated to average patch size ($r = 0.88$) and may primarily reflect this spatial attribute (correlations between number of patch types and these two spatial characteristics were both $r \leq 0.06$).

Our second spatial characteristic, the number of patch types appeared to influence separation success of plots by lichen community composition. Although we found no obvious relationship between number of patch types and separation success with the clustered community groups (community group Mantel's $r = 0.12$); variables with more groups, or patch types, tended to have greater separation success with raw community data (higher A-statistics; Fig. 1). Similarly, Goedickemeier et al. (1997) found reduced efficiency with stratified sampling when their number of strata was low. One of the outlying variables with a low A-statistic, yet many patch types, was geographic blocks. When we varied the number of patch types (by aggregating neighboring geographic blocks at varying scales), the positive relationship between number of patch types and A-statistics emerged again (Fig. 1a). Thus the varying levels of separation success of raw data
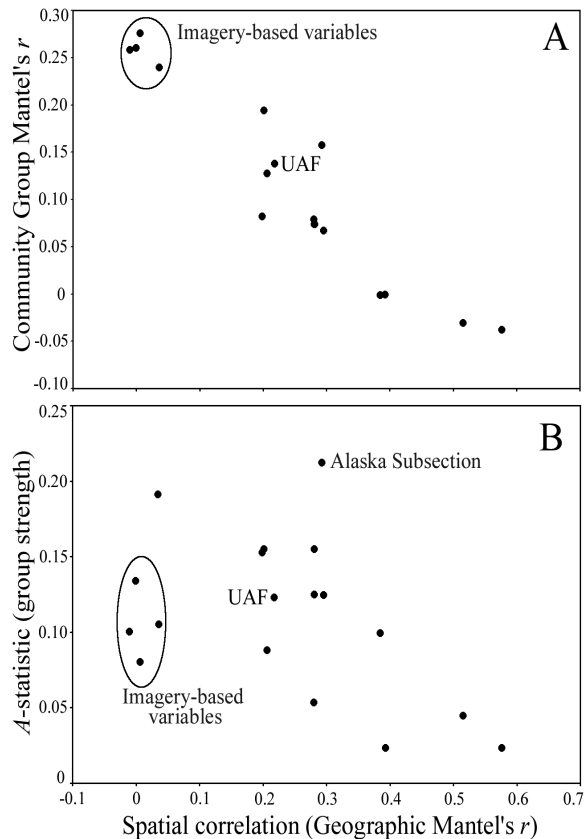
Fig. 2. Scatterplots of spatial correlation against two measures of separation success; (**A**) clustered community groups (community Mantel's r from comparisons of the seventeen GIS variables to the lichen community groups) and (**B**) raw community data (A-statistic from MRPP comparing lichen community composition among patch types). Each point represents one categorical variable. Both measures of separation success demonstrated negative relationships with spatial correlations ($r = -0.95$ community Mantel's r and spatial correlation and $r = -0.45$, A-statistics and spatial correlation). The latter relationship would strengthen following deletion of the four imagery-based variables with small patch sizes (circled in black; $r = -0.72$). The imagery-based variable with larger patch sizes is indicated as 'UAF. The categorical variable best balancing patch size and group strength (see Fig 1.) is labeled as 'Alaska Subsections'.

noted above (measured by A-statistics) reflects not only geologic or soil properties to which lichens respond, but also the number of groups used to divide the Preserve.

Average patch size, our third and final spatial characteristic, was also related to coarse patterns in lichen community composition. We found a strong negative relationship between community Mantel correlation statistics, relating lichen community types to the 17 GIS-derived categorical variables, and patch size ($r = -0.89$; Fig. 3a). Variables with smaller patches tended to have

stronger correlations with the three lichen community groups formed from cluster analysis. Turner et al. (1989) reported a similar loss of information with increasing grain size. We suggest this relationship in our analyses is due to two factors. First, implicit in small patch sizes is interspersion of different patch types, which reduces landscape contiguity. If we could map our lichen community groups, they would divide the Preserve not into three large contiguous patches, but rather many smaller discontinuous pieces of the three groups. The second factor that contributes to this negative relationship is the size of our sampling unit. Our plots were just slightly smaller than the average patch size of the four variables with the greatest congruence to the lichen community groups. Although we found a strong relationship of patch size with community group Mantel's r, we found no apparent pattern between average patch size and A-statistics, or the strength of each grouping ($r = -0.17$, with log transformed patch size; Fig. 3b).

This relationship between patch size and grouping strength based on raw community data (i.e., A-statistic) would be more strongly negative ($r = -0.49$) if the same four imagery-based categorical variables noted above were deleted. These four variables performed better than all others when compared to the clustered community data (see four highest data points in Fig. 3a), yet only moderately well when compared to the raw dataset (Fig. 3b). Unusually large community group Mantel's correlations resulted from 38% of the plots sharing a common patch type. Three of these four variables contained a single stratum (*Upland Dwarf Birch-Tussock Tundra* from Aggregate Ecological Types, *Upland Moist Dwarf Birch-Tussock Shrub* from Map Ecological Types and *Low Mixed Shrub-Tussock Tundra* from Map Vegetation Types) which strongly overlapped with a single lichen community group (*Lowland Group*; see Holt et al. 2007). This same lichen community group was also highly shared with two patch types from the fourth variable, Cover Types (23 % with *Mesic Dry Herbaceous* and 21% with *Open Low Shrub*; see Holt et al. 2007). Heterogeneity within this lichen community group, however, may have diminished the separating power of these imagery-based variables with the raw data; hence, these four variables achieved only moderate A-statistics.

Use as stratifying variables

The relationship between number of GIS-based patch types and separation of communities provides a basis for evaluating potential stratification principles to suit a budget, timeline or landscape area. The non-linear regression of A-statistics from 20 and higher-level groupings from cluster analysis represents target values for success of separating lichen community composition
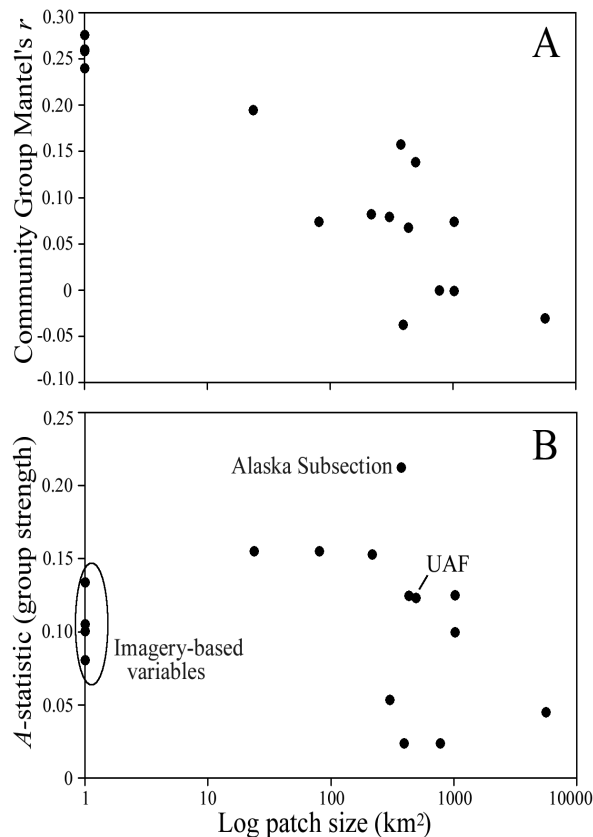
Fig. 3. Scatterplots of average patch size against separation success of (**A**) clustered community groups (community Mantel's r from comparisons of the seventeen GIS variables to the lichen community groups) and (**B**) raw community data (A-statistic from MRPP comparing lichen community composition among patch types). A general negative relationship ($r = -0.89$) between community Mantel's r and patch size demonstrates that fine-grained variation in community composition is not captured by coarse-grained patch characteristics. Alternatively, there appears to be little to no relationship between A-statistics and patch size ($r = -0.17$), unless the four imagery-based variables with small patch sizes (circled in black) were deleted ($r = -0.49$). The imagery-based variable with larger patch sizes is indicated as 'UAF'. The categorical variable best balancing patch size and group strength (see Fig 1.) is labeled as 'Alaska Subsections'.

by other variables, at various numbers of patch types: $y = 0.345 - 0.202e^{-0.188\,(x-2)}$; Fig. 1a. Separation success of other variables can then be evaluated relative to this curve or family of curves (Fig. 1b). Shaded bands, delineated by the non-linear regression curve and other curves with exactly a third the slope and intercept, show degrees of reduction from this target level of success. The only categorical variable included in the most successful band, the lightest color, was Alaska Subsections.

We were not surprised that Alaska Subsections, a variable based on climate, terrain, soils, vegetation and location, best discriminates lichen plots. These habitat factors strongly reflect the major elements, rockiness, substrate and topography, we previously found related to lichen community structure (Holt et al. 2007). However, several other categorical variables also based on similar factors (e.g. Aggregate Ecological Types, Map Ecological Types, Map Vegetation Types) did not perform as well. The exceptional success of Alaska Subsections, therefore, is partly due to its large number of groups yet moderate patch size. Alaska Sections, a more highly aggregated variable based on the same habitat information, however, was far less successful at separating plots by lichen community composition. Alaska Subsections, therefore, effectively balances complexity and simplicity to best capture fine-scale patterns in lichen community composition.

The next darker band, indicating variables moderately successful in discriminating lichen communities, includes over half the GIS-based categorical variables (Fig. 1b). Within this band, Cover Type had a moderately high A-statistic – suggesting its adequacy in separating lichen communities and justifying its use for our stratification. The Geographic Block variable used in our stratification also performed moderately well (Fig. 1b). As the only variable with arbitrary geographic boundaries and no ecological inputs, its low to moderate success in discriminating lichen communities was expected. Geographic blocks are, however, useful in providing equal representation from all parts of the study area. The four variables included in the weak success band, the darkest band, had few patch types. Among these four, Watersheds and Reindeer Ownership reflected minimal ecological information, thus their poor success was anticipated.

The value of this study is partially in its utility for future sampling stratifications. Our scope of inference, however, limits our results to the four cover types with high to moderate lichen cover that we sampled. If instead we had sampled every cover type present within the Preserve, the lichen communities would been even more disparate than our findings suggest (due to stark differences between the lichen-dominated sites and lichen-poor areas). These differences then might potentially alter our signal indicating which variables better separate plots by lichen community composition.

Based on our results, future lichen studies in neighboring Arctic regions would benefit by using the Alaska Subsection variable as sampling stratification. Considerable investment in more complicated stratifications, including imagery-based variables, likely sway land managers and other scientists towards using these instead. Our study, however, demonstrates the potential success of using a simple, relatively coarse-grained approach. Undoubtedly, high-technology, expensive, imagery-based variables have tremendous ecological value; however,

low-technology classifications also provide an effective, economical approach. Regional subsections strike an effective balance between patch size and degree of homogeneity of communities within strata.

Implications for other geographic areas and taxa

Our study not only provides suggestions specifically for future lichen sampling in northwestern Alaska, but also provides insight into stratified sampling as a general practice in ecological studies. As has been suggested by others (e.g., Ferrier & Smith 1990, Goedickemeier et al. 1997; Jenerette & Wu 2000), GIS is a vital tool for ecological survey design. We found, however, that GIS-based categorical variables varied in their ability to separate lichen communities. Those that best distinguished lichen communities related to substrate (e.g., Surficial Geology, Soil Units and Soil Series), yet this may differ for other taxa. Furthermore, we found, unsurprisingly, that each of the 17 variables we used, regardless of ecological relevance, better related to lichen community composition than did random groupings (Table 2).

Perhaps our most significant finding, however, was that scale matters. As evidenced by our results, every ecologist, regardless of specialization or region, who uses GIS data to stratify their study, should not select variables blindly. In addition to considering underlying habitat factors, stratifying variables should also be evaluated based the number of patch types and average patch size. Although we found a positive relationship between degree of community separation and number of groups; sampling across a large number of strata can be logistically difficult, time consuming and potentially expensive. As a general rule, we therefore recommend future studies use stratifying variables that balance the largest number of patch types while maintaining high landscape contiguity through moderately sized patches. Our study demonstrates categorical variables with smaller patch sizes tended to more closely resemble our ideal stratification of lichen community groups. The optimal balance between these two considerations likely varies with geographic area and taxa; however, our study underscores the importance of considering scale in stratification. In sum, we agree with Wiens (1989) that the spatial characteristics of a variable, including both number of patch types and patch size, can influence pattern detection and in turn can impact sampling efficiency.

References

Bond, D. & Devine, P. 1991. The role of geographic information systems in survey analysis. *The Statistician* 40: 209-215.

Brus, D.J. 1994. Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation. *Geoderma* 62: 233-246.

Cochran, W.G. 1977. *Sampling Techniques*. 3rd. ed. John Wiley and Sons, New York, NY, US.

Ferrier, S. & Smith, A.P. 1990. Using geographical information systems for biological survey design, analysis and extrapolation. *Australian Biologist* 3: 105-116.

Goedickemeier, I., Wildi, O. & Kienast, F. 1997. Sampling for vegetation survey: some properties of a GIS-based stratification compared to other statistical sampling methods. *Coenoses* 12: 43-50.

Gustafson, E.J. 1998. Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems* 1: 143-156.

Holt, E.A., McCune, B. & Neitlich, P. 2007. Successional and community gradients of arctic macrolichens and their relation to substrate, topography and rockiness. *Pacific Northwest Fungi* 2: 1-21.

Jenerette, G.D. & Wu, J. 2000. On the definitions of scale. *Bulletin of the Ecological Society of America* 81: 104-105.

Jongman, R.H.G., Bunce, R.G.H., Metzger, M.J., Mücher, C.A., Howard, D.C. & Mateus, V.L. 2006. Objectives and applications of a statistical environmental stratification of Europe. *Landscape Ecology* 21: 409-419.

Jorgenson, M.T., Roth, J.E., Emers, M., Davis, W., Schlentner, S.F. & Macander, M.J. 2004. *Landcover mapping for Bering Land Bridge National Preserve and Cape Krusenstern National Monument, Northwestern Alaska*. ABR, Inc. Environmental Research & Services, Fairbanks, AK, US.

Kernan, W.N., Viscoli, C.M., Makuch, R.W., Brass, L.M. & Horwitz, R.I. 1999. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology* 52: 19-26.

Knollová, I., Chytrý, M., Tichý, L. & Hájek, O. 2005. Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *Journal of Vegetation Science* 16: 479-486.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659-1673.

Legendre, P. & Fortin, M.-J. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80: 107-138.

Legendre, P. & Legendre, L. 1998. *Numerical Ecology*. 2nd. English ed. Elsevier, Amsterdam, NL.

Li, H. & Reynolds, J.F. 1994. A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology* 75: 2446-2455.

Mantel, N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Research* 27: 209-220.

Markon, C.J. & Wesser, S.D. 1997. *The Bering Land Bridge National Preserve Land Cover Map and its Comparability with 1995 Field Conditions*. US Geological Survey Open File Report 97-103, Anchorage, AK, US.

McCune, B. & Grace, J.B. 2002. *Analysis of ecological communities*. MjM Software, Gleneden Beach, OR, US.

McCune, B. & Mefford, M.J. 2005. *Multivariate analysis on the PC-ORD system*. Version 5. MjM Software, Gleneden Beach, OR, US.

Meisel, J.E. & Turner, M.G. 1998. Scale detection in real and artificial landscapes using semivariance analysis. *Landscape Ecology* 13: 347-362.

Mielke, P.W., Jr. 1984. Meteorological applications of permutation techniques based on distance functions. In: Krishaniah, P.R. & Sen, P.K. (eds.) *Handbook of statistics*, Vol. 4, pp. 813-830, Elsevier Science Publishers, Amsterdam, NL.

Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A., Liggett, W., Rathbun, S., Reckhow, K.H. & Young, L.J. 1999. Statistical issues for monitoring ecological and natural resources in the United States. *Environmental Monitoring Assessment* 54: 1-45.

Smartt, P.F.M. & Grainger, J.E.A. 1974. Sampling for vegetation survey: some aspects of the behavior of unrestricted, restricted and stratified techniques. *Journal of Biogeography* 1: 193-206.

Smouse, P.E., Long, J.C. & Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* 35: 627-632.

Till, A.B. & Dumoulin, J.A. 1994. Geology of Seward Peninsula and Saint Lawrence Island. In: Plafker, G. & Berg, H.C. (eds.) *The geology of North America*, pp. 141-152. The Geological Society of America, Boulder, CO, US.

Turner, M.G., O'Neill, R.V., Gardner, R.H. & Milne, B.T. 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology* 3: 153-162.

Van Patten, D.J. 1990. *Soil investigation of Seward Peninsula Area, Alaska*. USDA-Natural Resource Conservation Service in cooperation with Reindeer Herders Association, OCLC No. ocm36843275, Palmer, AK, US.

Viereck, L.A., Dyrness, C.T., Batten, A.R. & Wenzlick, K.J. 1992. *The Alaska Vegetation Classification*. USDA-Forest Service PNW-General Technical Report 286, Portland, OR, US.

Wiens, J.A. 1989. Spatial scaling in ecology. *Functional Ecology* 3: 385-397.

Wu, J. 2004. Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology* 19: 125-138.

App. 1. Correlation among all 17 GIS-based categorical variables, based on Mantel tests. Number along top row correspond to the numbered variables in the first column. Bold text indicates correlations stronger than ± 0.5.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Aggregate Ecological Types																
2 Alaska Ecosystems	0.08															
3 Alaska Sections	– 0.004	0.16														
4 Alaska Subsections	0.09	0.10	**0.56**													
5 Cover Types	0.20	0.04	0.01	0.08												
6 Ecoregions	0.01	0.26	0.41	0.23	0.06											
7 Elevation	0.02	0.14	0.12	0.12	0.09	0.21										
8 Geographic Blocks	-0.01	0.08	0.19	0.17	-0.07	0.11	0.05									
9 Map Ecological Types	**0.86**	0.06	0.01	0.13	0.14	-0.05	0.05	– 0.02								
10 Map Vegetation Types	**0.90**	0.06	– 0.004	0.11	0.16	-0.06	0.04	– 0.02	**0.96**							
11 Physiography	0.07	0.12	**0.75**	**0.62**	0.07	0.25	0.08	0.16	0.07	0.07						
12 Reindeer Owner	– 0.02	0.12	0.10	0.15	0.02	0.39	0.08	0.15	– 0.05	– 0.04	0.06					
13 Soil Series	0.06	0.11	0.41	0.41	0.10	0.23	0.14	0.08	0.06	0.07	0.36	0.07				
14 Soil Units	0.04	0.10	0.45	**0.51**	0.06	0.21	0.13	0.12	0.06	0.06	0.47	0.10	**0.81**			
15 Surficial Geology	0.13	0.12	0.36	0.34	0.05	0.07	0.10	0.18	0.18	0.16	0.43	0.06	0.32	0.33		
16 UAF Vegetation	0.20	0.19	0.31	0.36	0.07	0.08	0.13	0.12	0.22	0.22	0.36	0.03	0.30	0.32	0.24	
17 Watersheds	0.02	0.08	0.29	0.14	0.002	0.14	0.05	0.16	0.02	0.02	0.22	0.20	0.10	0.13	0.16	0.17

OPULUSPRESS